



Fantasy Football

Predicting Player Performance

Why?

- Available player score predictions are not always reliable.
- Many participants rely on an unscientific approach to player selection.
- Would be useful to individual players and leagues.

Note - This model uses ESPN's standard fantasy football scoring format.

The Training Data - Observations

- Each observation is a player's statistics from a specific game.

Total observations by position:

Wide Receivers



84,038

Running Backs



80,325

Tight Ends



49,318

Quarterbacks



31,031

Kickers



22,050

Data Wrangling and Cleaning

❖ Two sources of data

Player Stats

- Player game stats joined with player profile info
- Non-fantasy football positions filtered out (rows)
- Non-fantasy football stats filtered out (columns)
- Change team abbreviations to full name
- Remove stat columns with too many NaNs

Opponent Stats

- Deal with bye weeks by creating a “true” game of season column
- Remove stat columns with too much missing data
- Impute nulls where appropriate
- Parse dates

Feature Engineering

Player Stats

- Player stats converted to a mean of the previous three weeks*

Opponent Stats

- Convert stat columns from a weekly metric to a cumulative average for the season
- Winning percentage

*Three weeks is somewhat arbitrary. Further testing can determine the best window of time.

Choosing a Model: Random Forest

1st Iteration:

→ Unable to test deeper trees due to computing limitations

Hyperparameters:

- Max depth: 3, 4, 5, 6, 7
- N estimators: 10, 100

2nd Iteration:

- Max depth: 8, 9, 10, 11, 12
- N estimators: 10, 50, 100

Random Forest Performance

1st Iteration

- MAE: 3.74

Drawbacks

- Time required to test hyperparameters

2nd Iteration

- MAE: 3.64

Choosing a Model: Decision Tree

- More computationally efficient
 - Tree Depths Tested: 5-50
 - Results:
- ❖ Not the best performing (prone to overfitting), but best overall, considering the size of data

MAE: 3.73

RMSE: 5.34

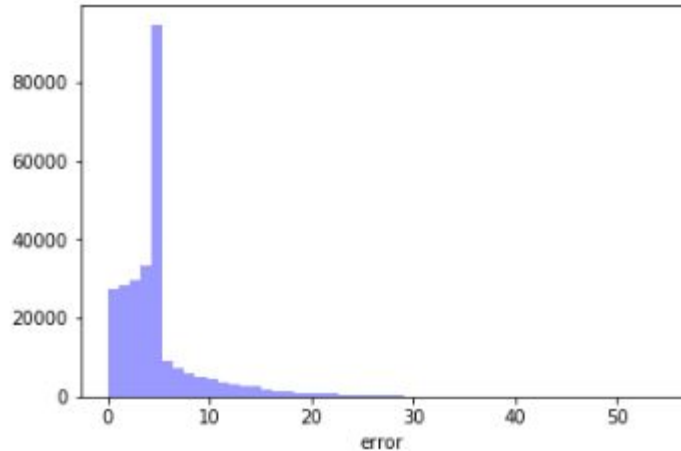
R-squared: 0.41

*Similar to the performance of the limited random forest that was tested previously.

Results

- Decision Tree Accuracy: 82% of predictions within 6 points of actual scores
- Most errors hover around 5 pts.

❖ Random Forest with tuned hyperparameters would likely improve predictions.



Future Optimizations

- Test different windows for player stat averages (5 weeks? whole season?)
- Bring in features initially left out of the model (height)
- Create a separate model for each position
- Add weights to more recent players